

Europäisches Patentamt
European Patent Office
Office européen des brevets



(11) **EP 1 072 984 A2**

(12) **EUROPEAN PATENT APPLICATION**

(43) Date of publication:
31.01.2001 Bulletin 2001/05

(51) Int. Cl.⁷: **G06F 17/30**

(21) Application number: **00305822.9**

(22) Date of filing: **10.07.2000**

(84) Designated Contracting States:
**AT BE CH CY DE DK ES FI FR GB GR IE IT LI LU
MC NL PT SE**
Designated Extension States:
AL LT LV MK RO SI

(30) Priority: **28.07.1999 CN 99110540**

(71) Applicant:
**International Business Machines Corporation
Armonk, NY 10504 (US)**

(72) Inventors:
• **Pan, Yue**
Winchester, Hampshire SO21 2JN (GB)
• **Yang, Li Ping**
Winchester, Hampshire SO21 2JN (GB)
• **Robertson, Lindon**
Winchester, Hampshire SO21 2JN (GB)

(74) Representative:
Burt, Roger James, Dr.
IBM United Kingdom Limited
Intellectual Property Department
Hursley Park
Winchester Hampshire SO21 2JN (GB)

(54) **Method and system for providing native language query service**

(57) The present invention provides a method for providing native language query service for Internet users by using a plurality of search engines, said method characterized by comprising steps of: (a) receiving at a site an original query request from one user; (b) selecting a suitable search engine; (c) translating said query words of native language into query

words of dedicated language of said selected search engine; (d) constructing a new query request directing to said search engine; (e) sending said new query request and receiving a returned query result; (f) sending said query result back to said user as a query result in relation to said original query request.

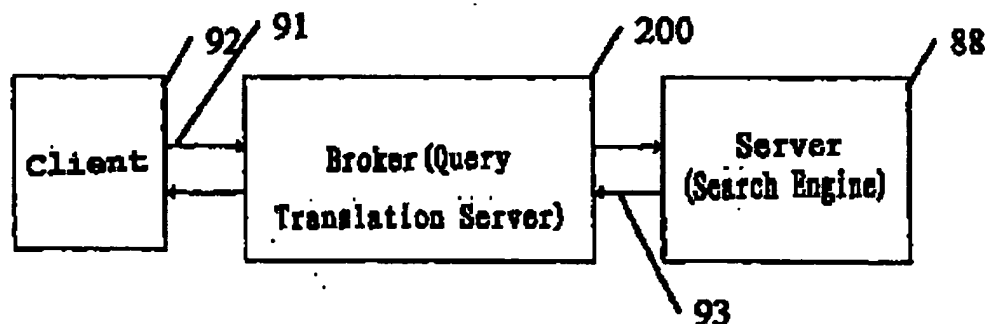


FIG. 3

Description

[0001] This invention relates to a method and system of information retrieval, especially relates to a method and system for providing query service for Internet or database users, and particularly relates to a method and system for providing query service in native language for Internet or database users.

[0002] With the development of computerized information source such as remote networks, users of data processing system are able to connect to other servers and networks so as to retrieve massive electronic information which can not be accessed in traditional electronic media. This kind of information media is increasingly substituting traditional one for information distribution such as newspaper and even television.

[0003] In communication, different computer networks are connected with each other through "gateway" which handles the data transmission and message transformation from sending network to receiving network. Gateway is a device used to connect different networks (that use different communication protocols) so that information can be transmitted from one network to another. Gateway also transforms the format of information so that it is compatible with the protocol adopted by the other network.

[0004] Internet is one of the remote networks generally used today. Internet originates from ARPAnet of the United States and now is the largest open computer network consisting of numerous networks connected with each other. Internet uses TCP/IP protocol that is the abbreviation for "Transport Control Protocol/Internet Protocol". The software protocol was developed by Defense Department of the United States for communication between computers. Internet can be described as the distributed computer network consisting of computers that connect with each other and run network protocols that enable users to share network information. Because of the requirement to share information in general use, remote networks such as Internet are developed into "open" network. For open system, developers can design software application for specification operation or service almost without any restriction. For details about node, object and link of Internet, readers can refer to textbook "Mastering the Internet" by G.H.Cady et al and published by Sybex Company, Alameda, California, USA.

[0005] The electronic information exchanged among computers is usually formatted in hypertext. Hypertext is a prior technology that couples text, image, audio and action into a complex, nonlinear network and permit users to "browse" or "surf" by relevant title. The term "hypertext" is created in 1960s to refer to the computer files which show distinct nonlinear structure from linear structure in book, film and language.

[0006] On the other hand, the term "hypermedia" emerging recently is a synonym of "hypertext" but emphasizing on the non-text constituents of hypertext such as, animation, audio, video and etc. Hypermedia image, audio, video or their arbitrary combination can be synthesized into information storage and retrieval system. The hypermedia and hypertext, especially the interactive ones selectable and controllable by the users, are constructed in an environment, where the work and study are parallel to human thoughts. That is, they permit the users to establish correlation between the topics instead of moving sequentially from one motion to the next one. The hypermedia and hypertext are linked in such a way that the users can jump from one topic to another relevant one during search period. Hyperlinks are embedded in hypermedia or hypertext files and lead to other hypermedia or hypertext files when "clicked" by the reader.

[0007] World Wide Web(WWW), or "Web", is the multimedia information retrieval system on Internet. Web is the most common method of transferring data in network. Other methods include FTP (File Transfer Protocol) and Gopher, etc., but they are not so common as web. On web, clients access the service provided by web server using Hypertext Transfer Protocol (HTTP), where HTTP is a well-known application protocol, which permits the users to use a standard page description language called Hypertext Markup Language (HTML) to access various kinds of files (e.g. text files, graphics, images, audios, videos, etc). HTML provides underlying file formats, and allows developers to specify the links to other servers and files. Link that denotes the network path to other web server is specified in Universal Resource Locator (URL). There is dedicated syntax for specifying network path by URL.

[0008] URL is typically formatted as: http://somehost/someprogram?parameters... in which "somehost" is the host address of the URL, "somedirectory" is the directory name where the web page indicated by the URL can be found. The usual way to decompose an URL into an actual address of a Web server is by using a domain server. In the internet or Intranet, a domain name server converts the host name into an actual Web address. One example of the domain name servers is the Domain Name Service (DNS) in Internet. The procedure, in which a Web user requests a host name and an address from the domain name server, is called parsing. In the TCP/IP, the domain name server parses the host name and makes one or more IP address lists, which are returned to the Web clients, who requested on the HTTP. Each IP address specifies a server, which is used to process the request contents sent by the browser.

[0009] WWW which employs hypertext protocol follows a client/server architecture. At the client side is the browser software that sends requests to web server and then explains, displays or plays the hypertext information and other multimedia files returned by a web server.

[0010] While thousands upon thousands of companies, universities, government offices, museums and municipal authorities publish their homepages on the web, Internet grows to be a very valuable information source. Even a beginner is able to browse thousands of web pages or news groups after a little practice. Hence the Internet accesses and

the corresponding markets are growing rapidly.

[0011] Today World Wide Web holds huge amount of information which is organized by hypertext pages. The number of pages is so large that it is impossible for people to follow site by site, link by link even within one site like IBM's, let alone the whole web. Thus it is very important for users to efficiently, rapidly and conveniently find the interested information on world wide web.

[0012] Currently, there are many kinds of information retrieval facilities. One of them is a directory service. In a directory service, web sites/pages are classified into categories and each category is further classified into sub-categories. Further more, categories are organized hierarchically and presented as web pages with links to sub-categories. By exploring layer by layer, finally the users may discover a set of relevant web sites/pages eventually. With the amount of web sites/pages increasing explosively, directories grow to huge sizes. For example, Yahoo! currently has 25,000 categories that hold over half a million sites.

[0013] Another kind of information retrieval facility is a search engine. A search engine indexes web sites/pages by using key words. A user inputs query words, that are key words acceptable by the search engine, and then he gets related web sites/pages. More and more people would rather query search engines for web sites/pages by using key words than exploring layer by layer in the directories. Category entries may also be indexed for searching. For example, in Yahoo! both category entries and web sites are presented as search results.

[0014] Computer application software such as free-of-charge or low cost Internet search engines let people easily get to the Web sites by querying so as to obtain corresponding topic information from those sites. When users are surfing on the Internet, directories and search engines help them so much that they are recognized as "portal sites" to the Internet. Famous directories/search engines include Yahoo!, AltaVista, Excite, etc. Sina (www.sina.com.cn) is a very famous Chinese portal site. However, most of web pages are written in English and also cataloged/indexed in English. Users who can not read English can not enjoy the convenience.

[0015] To solve this problem, there are some web page translation software that can statically or dynamically translate web pages from one language into another, such as from English into Chinese, or from German into English, etc. Some software also reserves HTML tag structure such that users can follow links marked in their native language without difficulty. Moreover, users can also utilize the directories since the names of categories are also dynamically translated.

[0016] Even so, Internet users still can not query and retrieve in their native language different from that search engines used in indexing. The reason is that the translation of the query words from the native language into other languages that search engines are using is lacking, even if web page translation software does exist.

[0017] Some query translation prototypes/solutions partly solve the problem. They provide predefined forms to collect input from the users and then translate the user input into a language, that the search engines can accept, and construct query commands, which are later sent to particular search engines. However, they have the following drawbacks:

[0018] Firstly, many Internet search engines support a combinatorial query and/or an second query to narrow down the scope. In the first case, the word string to be matched as a search condition is logically combined with other search conditions such as categories or date of publication to be retrieved. In the second case, the second search is done by using the result returned by the first search. However, these functions are lost when users use these query translation software.

[0019] Secondly, ambiguity of translation cannot be properly handled. When query words are translated from one language into another, the mapping is often one-to-many. The problem is further complicated by the presence of synonym (a word with multiple meaning) in many languages especially in Chinese. For example, the Chinese word

“腔”

(PINYIN: qing1) has four meanings and each meaning may correspond to several English translations. The first meaning can be translated into English words "cavity" and "chamber". The second meaning can be translated into English word "speech". The third meaning can be translated into English words "tune" and "pitch". The fourth meaning can be translated into English words "accent" and "tone". There are many such kind of synonyms in Chinese.

[0020] The query translation server for Internet search engines which is described by this invention can eliminate the above drawbacks.

[0021] Thus, the first object of this invention is to provide a method for Internet user to utilize query service in the native language.

[0022] The second object of this invention is to provide a kind of system for Internet user to utilize query service in the native language.

[0023] The third object of this invention is to provide a kind of system for database user to utilize query service in the native language.

[0024] For the implementation of said first object, this invention is a method for providing native language query

service for Internet users using a plurality of existing Internet search engines, each search engine having a respective dedicated query language, adapted for receiving query request containing query terms in said dedicated language and returning query results in relation to said query terms, said method being characterized by comprising the steps of:

- 5 (a) receiving at a site an original query request from one user, in said original query request there are query terms of said user in his/her native language;
- (b) selecting a suitable search engine from said plurality of search engines;
- 10 (c) translating said query terms in the native language into query terms in the dedicated language of said selected search engine;
- (d) constructing a new query request directing to said search engine based on said original query request and said query terms in said dedicated language;
- 15 (e) sending said new query request to said selected search engine and receiving a returned query result;
- (f) sending said query result back to said user as a query result in relation to said original query request.

20 **[0025]** This invention is also a method for providing a native language query service for Internet users by using a plurality of existing Internet search engines, each search engine having a respective dedicated query language, adapted for receiving query request containing query terms in said dedicated language and returning query results in relation to said query terms, said method being characterized by comprising steps of:

- 25 (a) receiving at a site an original query request from said Internet user, said original query request containing an URL requested by said user, said URL having a prefix for designating said site;
- (b) removing said prefix from said URL;
- 30 (c) selecting a suitable search engine from said plurality of search engines;
- (d) if said removed prefix is a redirect prefix, performing the following steps of:
 - (d1) sending a request containing said URL to said selected search engine and receiving a web page as a response;
 - 35 (d2) adding a translation prefix in front of URLs that need query terms and adding redirect prefix in front of other URLs in said web page, so as to form a new web page;
- 40 (e) if said removed prefix is a translation prefix, performing the following steps of:
 - (e1) translating query terms of user's native language in parameters of said URL into query words in said dedicated language of said selected search engine;
 - 45 (e2) replacing query terms of user's native language in parameters of said URL with query terms in said dedicated language;
 - (e3) adding a redirect prefix in front of said URL;
 - 50 (e4) generating a new web page, embedding said URL and a script program in said web page, said script program enabling the client which receives said new web page to perform a step of automatically sending another original query request based on said URL embedded in said web page;
 - (f) sending said new web page back to said user as a query result in relation to said original query request.

55 **[0026]** For implementation of said second object, this invention provides a system for providing native language query service for Internet users using a plurality of existing Internet search engines, each search engine having a respective dedicated query language, adapted for receiving query request containing query terms in said dedicated lan-

guage and returning query results in relation to said query terms, said system characterized by further comprising a query translation server which comprises:

- 5 a client interface, for receiving query requests sent by clients and returning query results to said clients;
- a request dispatching apparatus, for receiving said query requests from said client interface, removing prefixes from requested URLs, and distributing said query requests to different components based on said removed prefixes;
- 10 a web page obtaining apparatus, for receiving a query request whose prefix is a redirect prefix from said request dispatching apparatus, sending said query request to a search engine designated by an URL and obtaining an URL requested web page;
- 15 a web page modifying apparatus, for forming a new web page by adding translation prefixes in front of URLs that need query terms and adding redirect prefixes in front of other URLs in the obtained web page, and sending said new web page to said client interface;
- 20 a query translation apparatus, for receiving a query request whose prefix is a translation prefix from said request dispatching apparatus, translating query terms in user's native language in the requested URL into and replacing them with query terms in a dedicated language of said search engine, and adding a redirect prefix in front of said URL;
- 25 a web page generation apparatus, for generating a new web page, embedding said URL and a script program in said web page, and sending said new web page to said client interface, said script program enabling a client which receives said new web page to perform a step of automatically sending another query request based on said URL embedded in said web page.

[0027] For implementation of said third object, this invention describes a system for providing native language query service for database users using a plurality of databases, each database having a respective dedicated query language, adapted for receiving query request containing query terms in said dedicated language and returning query results in relation to said query terms, said system characterized by further comprising a query translation server which comprises:

- 35 a client interface, for receiving query requests sent by clients and returning query results to said clients;
- a query translation apparatus, for receiving query terms in user's native language in the query requests received by said client interface into and replacing them with query terms in the dedicated language of the database;
- 40 a query result obtaining apparatus, for sending the translated query requests to the databases designated by said query requests and obtaining query results.

[0028] By using the method and system of this invention, a user need only to select one of Internet search engines and use its built-in complex query functions to query in his/her native language without considering the language used by the search engine. Thus users from other countries can share the great convenience provided when they retrieve or query for information in Internet by using their native languages.

[0029] Some search engines provide built-in complex query functions including combinatorial queries, evolutionary queries, and restricted queries. With the method and system of this invention, the above functions are retained.

[0030] According to the method and system of this invention, it is also possible to do site-specific query translation and discriminate multiple meanings of a synonym and multiple translations of a meaning.

50 [0031] To enable the above features, it is not necessary to modify any client software such as a browser or the search engines. Thus the method and system of this invention is easy to implement and promote.

[0032] With reference to the accompanying figures that describe the preferred embodiment of this invention, the features and advantages of this invention will be more obvious.

55 Figure 1 is the client/server architecture that is widely applied in prior art.

Figure 2 is the computer network according to the preferred embodiment of this invention.

Figure 3 is the client/broker/server architecture according to the preferred embodiment of this invention.

Figure 4 is the flow chart of execution of the first embodiment of this invention that provides query service in native language for Internet users.

5

Figure 5 is the flow chart of execution of the second embodiment of this invention that provides query service in native language for Internet users.

10

Figure 6 is a schematic diagram illustrating a system of this invention that provides query service in native language for Internet users.

Figure 7 to 10 are the demonstrations of a browser when a user queries in Chinese using the method and system of this invention.

15 [0033] Figure 1 is the client/server architecture that is widely applied in the prior art. Client request (such as query request, news request) 91 is sent to server 88 by client 92. Server 88 may be a remote computer system that can be accessed through Internet or other communication networks. For example, it is possible to run client 92 on a user personal computer (not shown) while running an Internet search engine on server 88.

[0034] Per client request, server 88 scans and searches for original information (such as online news or news group), then returns the filtered electronic information to client 92 as a server's response 93.

20 [0035] Client 92 may run on the first computer, the server process may run on the second computer system. They communicate with each other through the communication media, thus providing distributed service and enabling multiple clients to utilize the information collection ability of one server.

[0036] On world wide web, it is usually possible to retrieve information by means of browser compatible with HTML running on client computer. When the user a of browser specifies a route by URL, client computer sends a request to domain name server so as to convert the host name in URL into IP address of the server. Domain name server responses with a list of one or multiple IP addresses. One of these IP addresses is used by the browser to establish the connection with a server. If the server is available, it returns the documents or other objects formatted according to HTML. On the server is running the corresponding server software, which provides information in the form of HTTP response to the client. HTTP response corresponds to the Web pages constructed in HTML language or data produced by other servers. Web browsers have become the main interface for accessing many networks and servers.

30 [0037] For example, when a user retrieves information using search engines, he/she first runs one kind of browser software on his/her personal computer and then designates and inputs an URL for a search engine, such as the URL of Yahoo! site, "http://www.yahoo.com". Yahoo! server returns its homepage to the browser. The homepage is formatted in HTML and the browser displays the content according to HTML specification. The user can click on the hypertext using pointing device or input query term for retrieval or query.

[0038] The above browser can be any kind of browser available on the market, such as Navigator and Communicator from Netscape, Internet Explorer from Microsoft, Mosaic and Lynx from National Center for Super Computer Application (NCSA) in Urbana-Champaign of Illinois. Any other browsers which can provide functions specified by HTTP can be used. Web browser is typically a program that is able to parse and display HTML documents, though those skilled in the art will appreciate that the future browsers will be able to handle other markup languages such as dynamic HTML and XML documents.

45 [0039] A typical server includes an IBM RISC/6000 computer (based on so-called Reduced Instructions Set Computer) running an AIX (Advanced Interactive Executable program version 4.1 or higher version) operating system and server program. The server accepts requests from the client through dial-up computer network in order to obtain the files or objects requested by the client or to execute the requested tasks. Various types of RISC-based computer are described in many publications of IBM Corporation, such as (Hardware Technical Reference Manual of RS/6000, 7013 and 7016 POWERstation and POWERserver) (order number SA23-2644-00). The AIX operating system is described in the first edition of (Technical Reference Manual of AIX operating system) published by IBM Corporation in November 1985 and other publications. While the above-mentioned are practical, any suitable hardware/operating system/server combination can be employed. For example, the server architecture of WINDOWS 95/98/NT of Microsoft Corporation is available.

55 [0040] Figure 2 illustrates the computer network 80 following client/server architecture. Internet is a well-known computer network following the architecture. However, those skilled in the art should understand that the computer network 80 can be implemented not only by Internet but also by other distributed computer network such as Intranet.

[0041] Theoretically, Internet is a big network consisting of server 88 which can be accessed by a client, typically a personal computer, through some Internet service provider 84, such as Internet America, or online service provider, such as America Online, Prodigy, CompuServe, etc. Each client can run a browser (a client) to access server 88

through access providers. Every web server 88 is a web site. In this invention, web site may be a site providing directory service, or a search engine, or a system coupled with server executing the method of this invention and multiple search engines to provide native language query service for Internet users.

[0042] Since this invention relates to the network transmission, it is helpful to have an understanding of network and its operating principle. There are not details of the network to which this invention can be applied. For example, this invention can be applied to a global network such as Internet.

[0043] Based on the client/server architecture illustrated in Figure 1, this invention adopts a client/broker/server architecture for Internet illustrated in Figure 3. Actually, the above client/server architecture exists between client 92 and broker 200 as well as between broker 200 and server 88. In Figure 3, client 92 first sends request 91 to broker 200. Broker 200 processes the request 91 and sends the processed request to server 88. Server 88 provides corresponding service per the request and return response 93 to broker 200. Broker 200 processes response 93 and returns the processed response to client 92.

[0044] The site executing the method of this invention acts as broker 200. Client 92 may be any existing browser, server 88 may be any existing site, which provides search engines.

[0045] Currently, every search engine uses a dedicated language. One search engine can accept only query terms in that dedicated language. For example, Yahoo! can accept query terms in English, thus its dedicated language is English. Another example is Sina (www.sina.com.cn) whose dedicated language is Chinese. The native language of user may be any language that is the same as or different from the dedicated language of the search engine.

[0046] Figure 4 is the flow chart of execution of the first embodiment of this invention that provides query service in native language for Internet users. The method utilizes existing multiple search engines on Internet, each one using dedicated language and accepting query request containing query terms in the dedicated language and returning query results relevant to the query terms.

[0047] At step 401, the original query request is accepted which contains the query term in the Internet user's native language. At step 402, a suitable search engine is selected from multiple search engines per the request. At step 403, the query term in native language is translated to the specific language used by the selected search engine. At step 404, a new query request is constructed according to the original query request and the specific language used by the identified search engine. At step 405, the new query request is sent to the selected search engine and query result is accepted. At step 406, the query result is returned to the user as a relevant query result of the original query request.

[0048] One may develop a program implementing the query translation method of figure 4 by using Java language so that it can run on various operating systems including the above AIX, Windows 95/98/NT, etc. As this program code can be made compatible with Servlet API Standard, it can be embedded into many Web sites and can run with other Web servers. Of course, it is also possible to configure a new site in the network shown in Fig. 2, dedicated for executing the method of providing query services in native language to Internet users, as shown in Fig. 4. In such cases, the newly configured site is called a query translation server.

[0049] The implementation of each steps of Fig. 4 is described in details below.

[0050] Step 401 in figure 4 may be implemented in many ways.

[0051] In a simple way, the query translation server runs as a proxy server. This requires that the user configures the web client software so that all the query requests sent by the client software are sent to query translation server of this invention. This approach brings inconvenience to user sometimes and increases the burden of the query translation server and computer network since files that need not be translated such as image and audio files have to go through it too.

[0052] Another way is that the query translation server runs as a normal server but uses an URL transformation schema. The schema automatically adds prefixes including the network address and path of the query translation server in front of the requested URL when the user sends a query request. Therefore, all the query requests of the client are automatically transferred to the query translation server of this invention.

[0053] Especially, if the unique URL transformation scheme mentioned below is used, no modification of the client software is needed, thus greatly facilitating the operation of the users. Another improvement lies in that in order to decrease the burden of the query translation server and computer network, different prefixes are placed in front of different URLs. This point is easy to see by analyzing the query request sent by the client. Query requests may be classified into 2 types. One type contains query terms and if the query terms use a kind of language different from the specific language used by the search engine, the query terms must be translated into ones in the specific language before they are sent to the search engine. Another type does not contain any query terms and it is not necessary to translate them, so the query translation server can redirect directly to the corresponding search engine. Thus these two types of query request should be treated differently. In this invention, different prefixes are added to the URLs of these two types of query request, i.e., translation prefix and redirect prefix, so that these query requests are treated differently in the query translation server of this invention.

[0054] Without any configuration change of the client software, the process of adding prefixes is done on the query

translation server. The query translation server first gets the homepage of a search engine and adds appropriate prefixes to the URLs (i.e. the URL of the query translation server) embedded in the homepage and returns the thus modified homepage to the client. In the following queries, similar URL transformation is applied to every web page returned by the search engine.

- 5 [0055] Steps 402 and 403 in figure 4 can be implemented by using predefined templates of this invention for search engines. The templates may be stored in the memory or disk of query translation server and retrieved in a file or database. The templates contain necessary information for the search engines supported by the query translation server, the information is used at step 402 and step 403.

[0056] For example, the following templates for search engines may be stored in a template file.

10

yahoo=\

15

```
url=http://search.yahoo.com/bin/search,\
parameterName=p,\
language=English,\
assemblerClass=Generic,\
dictionaryContext=
```

20

ibm=\

25

```
url=http://www.ibm.com:8080/db2search/,\
parameterName=q;s,\
language=English,\
assemblerClass=Generic,\
dictionaryContext=ibm
```

30

...

35

- [0057] In the example, the first two templates are for Yahoo! and IBM respectively. Every template defines the URL for the search program of the search engine, a list of parameters to be translated in the query command, the specific language acceptable by the search engine, the rule for combination of query terms, and the dictionary context of specific site. The rule for combination of query terms is used to specify the relationship between query terms, for example, query terms may be combined by "+", "-", "AND", "OR", or "NOT". The rules may be different for different search engines. The dictionary context of a specific site is used to name the specific dictionary when the query is done in a specific site.

- 40 [0058] It should be indicated that using templates is not the only way to support the method of this invention. Other ways may be used to store and retrieve the necessary information for Internet search engines. This is the common sense of developers in this domain and thus do not impose any restriction to this invention.

[0059] Suppose the original query request over Yahoo! site is received at step 401 which contains the following URL:

50

http://search.yahoo.com/bin/search?p=游戏-66

55

[0060] At step 402, a template of search engine is looked up that matches the URL. The following search engine template is found:

yahoo=\

```
url=http://search.yahoo.com/bin/search,\
parameterName=p,\
language=English,\
assemblerClass=Generic,\
dictionaryContext=
```

[0061] This is a template for Yahoo!. Thus it is known that the user wants to query Yahoo! search engine. It is also known that parameter list in the URL is

"p=游戏",

since the template specifies that only parameter "p" needs to be translated. The specific language of the search engine is English. The combination rule for the search engine is named as Generic, which defines the rule of combination of query terms for Yahoo! search engine in the form of a function. The dictionary context is null, indicating that general dictionary should be used.

[0062] At step 403, the query term

"p=游戏"

in native language (Chinese here) of the parameter list

"游戏"

is translated into query terms in specific language (English here), for example "recreation".

[0063] The following has to be noted. If there is no matched search engine template in the template file, a default rule is to be used for that URL. The default rule first checks which one is the dedicated language used by the search engine designated by that URL (corresponding to "language" parameter in the template) based on the language employed by the Web page (just that page from which the user sends out the request) embedded into that URL, or locates from the URL history record processed by the query translation server an URL most similar to that URL, and checks the language used by the returned Web page. The checked result is labelled as the dedicated language. Then the location of the query term parameter (corresponding to "parameter Name" parameter in the template) is found based on the parameter of the query parameter list expressed in the native language. After the determination of the location of the query term parameter, the query term parameter in the native language is translated into a query term in the dedicated language. The parameters corresponding to "assemblerClass" parameter and "dictionaryContext" parameter in the template take their default values.

[0064] At step 404, new query request for the selected search engine is made from the original query request and the query term in the specific language used by the search engine. In this example, new query request contains the following URL:

<http://search.yahoo.com/bin/search?p=recreation&y=66>

[0065] Step 405 may be executed directly on query translation server. However, the simple approach may result in loss of some proprietary features of this invention, such as providing multiple translation choices to users. A better alternative is to predefine some script program that will be executed on the client when it is downloaded at the client as part of the responsive web page. So the web page is made from the query request coming at step 404 and the predefined script, then it is returned to the client as a response. It is the characteristic of the script that it will be executed automati-

ically when the response is received at the client.

[0066] Since Chinese term

5

“游戏”

may be translated into several English terms, such as "recreation", "game" or "play", steps 403 and 404 may generate several URL, such as:

10 http://search.yahoo.com/bin/search?p=recreation&y=66,
 http://search.yahoo.com/bin/search?p=game&y=66, and
 http://search.yahoo.com/bin/search?p=play&y=66.

[0067] One of them, for example

15

 http://search.yahoo.com/bin/search?p=recreation&y=66

is labeled as the default URL.

[0068] The multiple translation options produced during translation can be displayed to the user with the help of the
 20 written script program, and the default translated query request is sent to the search engine via the query translation
 server. The server adds a segment of script program to the Web page, which is then sent to the client. The script pro-
 gram itself provides a function such that when the client receives the web page, the script program embedded in the
 web page is automatically executed. This is well known to those skilled in the art and will no more be discussed.

[0069] Step 406 can also have many ways of implementation, which are similar to those at step 401. If step 401 is
 25 implemented by using an agent server, step 406 only simply returns the web page returned by the search engine to the
 client. If step 401 is implemented by using URL transformation, step 406 returns the web page to the client only after
 the above-mentioned URL transformation of the web page returned by the search engine.

[0070] Fig. 5 is the flow diagram of the second embodiment of the method of this invention for providing query serv-
 30 ice in the native language to Internet users. The method employs multiple existing search engines, each of which has
 its own dedicated language for receiving the query requests of query terms in its respective dedicated language and for
 returning the query results related to the query term.

[0071] Step 501 receives the original query request of the user. Assume that the query translation method shown
 in Fig. 5 operates on one portal site, and the portal site of the URL is, for example, http://portal_site/. Then, in the client
 request received at step 501, the requested URL will certainly contain http://portal_site/.

35 [0072] Step 502 removes the prefix of the requested URL. As explained below, the prefix may be:

 http://portal_site/redirect/(redirect prefix), or
 http://portal_site/translate/(translation prefix).

40 [0073] Of course, these two prefixes do not impose any restrictions on this invention. In practice, there may be
 many kinds of prefixes, depending on various functions, so far as these prefixes conform to the dedicated syntax.

[0074] Step 503 selects a suitable search engine from a plurality of search engines.

[0075] Step 504 judges if the removed prefix is a redirection prefix http://portal_site/redirect/or not. If the judgment
 45 of the step 504 is "No", the processing proceeds to step 505; otherwise it proceeds to step 509. As in this example there
 are only two prefixes, namely redirection prefix and translation prefix. Hence in the steps before step 505 the removed
 prefix is certainly a translation prefix, i.e. http://portal_site/translate/.

[0076] If in the request received at step 501 the requested URL is:

50 http://portal_site/translation/http://search.yahoo.
 com/bin/search?p=游戏&y=66,

the URL with prefix removed at step 505 will be

55

 http://search.yahoo.com/bin/search?p=游戏&y=66

[0077] In step 505, the above query term

5

“游戏”

of URL in the native language (in this example it is Chinese) is translated into a query term in the dedicated language (in this example it is English) in the similar way as steps 402 and 403 in Fig. 4 (i.e. by using search engine template). Then at step 508 the query term in the native language is replaced with those in the dedicated language.

10

[0078] As the term

“游戏”

15 in Chinese may be translated into many English terms, such as "recreation", "game" or "play".

[0079] Step 508 can produce a plurality of URLs, e.g.

20

http://search.yahoo.com/bin/search?p=recreation&y=66,
http://search.yahoo.com/bin/search?p=game&y=66, and
http://search.yahoo.com/bin/search?p=play&y=66,

[0080] One of the above three URLs, for example,

25

http://search.yahoo.com/bin/search?p=recreation&y=66 is labelled as a default URL.

[0081] To support the combined query requests, the combination operators such as "+", "-", "AND", "OR", or "NOT", in the query request are identified by means of the combination rule specified by the template of the search engine. Then the query terms in the native language separated by the combination operators are translated into query terms in the dedicated language.

30

[0082] To support the site-specific search, an appropriate dictionary is selected by the dictionary context (dictionary-Context) of template for the search engine and translation is done using the dictionary. For example, if the user query is over IBM site, the query term

35

“服务器”

may be translated into and substituted with "RS/6000", "AS/400" or "S/390". If the user query is over HP site, the query term

40

“服

45

务器”

may be translated into and substituted with "HP3000" or "NetServer", etc.

50

[0083] After step 508, at step 507, redirect prefix is added in front of the above URL. Thus, the following three URLs are formed:

55

http://portal_site/redirect/http://search.yahoo.com/bin/search?p=recreation&y=66,
http://portal_site/redirect/http://search.yahoo.com/bin/search?p=game&y=66, and
http://portal_site/redirect/http://search.yahoo.com/bin/search?p=play&y=66

[0084] Then the processing proceeds to step 508. At step 508, a new web page is generated in which the above three URLs and a segment of script are embedded. The script will be executed automatically by the client when the web page is received. By executing the script, the client displays multiple translation choices of Chinese query term

5

“游戏”

to the user. There are many ways to display the translation choices. A good way is to open a new window dynamically and show the translation choices in that window. Every translation choice has a hyper-link, that is, the URL contains the URL of translation. In the example, Chinese query term

10

“游戏”

has three translation choices in English: "recreation", "game" and "play". These translation choices are shown in the new window. The user can click on either of the three translations using pointing device and the browser then sends another original query request. Another function of the above script is to send a new request based on the default URL (http://portal_site/redirect/http://search.yahoo.com/bin/search?p=recreation&y=66).

[0085] A special segment of script can be embedded in the web page at step 508 to provide function with which a user can modify the template of search engine. For example, it may display the content of templates of search engine such as the four items described above. User may modify the content of templates and submit the modification to query translation server.

[0086] The processing proceeds to step 511 after step 508. At step 511, the newly generated web page is returned to client as a response. As described above, if the result of judgment at step 504 is "YES", then the process goes to step 509. At step 509, query translation server retrieves the web page identified by the URL from Internet.

[0087] For example, if the web page contains http://portal_site/redirect/http://search.yahoo.com/bin/search?p=recreation&y=66 is returned to client at step 511, then the client executes the script described above automatically and sends new HTTP request. After the processing at steps 501, 502, 503 and 504, the web page identified by following URL is retrieved at step 509:

30

<http://search.yahoo.com/bin/search?p=recreation&y=66>

[0088] This is the detail of the process. The request is sent to Yahoo! search engine. Yahoo! search engine searches for the query terms and returns the search results by web page. Then the process goes to step 510.

[0089] At step 510, the web page received from search engine is parsed and the embedded URL is transformed.

[0090] This is a simple introduction to hypertext markup language (HTML) used in web page formation. HTML uses tags or tag pair quoted in ϕ to control the display of web page content. Every HTML file must begin with ϕ HTML ϕ tag and ended with ϕ HTML ϕ tag. Other tags include: (A HREF) and (A) for anchor and title of link, (FRAME) and (FRAME) for frame content, (FRAME SRC) for introducing images into the frame, (FORM ACTION) for relating to a form. Tags can include other tags to become a nested one, thus creating enhanced objects. For example, (A HREF) tag may include an (IMG SRC) tag to create a linkable image. Of course, the above tags can only be regarded as examples since HTML is an evolving language.

[0091] At the above step 510, three types of URL in web page are transformed. They are,

- (1) "href" parameter of (A) or (AREA) tag
- (2) "src" parameter of (FRAME) tag.
- (3) "action" parameter of (FORM) tag.

[0092] Redirect prefix are added in front of type (1) and type (2) URLs

50

http://portal_site/redirect/

Translate prefix are added before type (3) URL

http://portal_site/translate/

[0093] After transforming the URL embedded in the web page, a new web page is generated. Then the process goes to step 511. At step 511, the newly generated web page is returned to client as a response.

[0094] Figure 6 illustrates the system of this invention that provides native language query service to Internet users. Included in the dotted frame is the query translation server of this invention which couples with search engines in Inter-

net 600 to make up the system of this invention that provides native language query service to Internet users. As shown in Figure 6, the query translation server of this invention consists of client interface 601 for receiving query request from client and returning query result to the client; request dispatching device 602 for checking the type of prefix of URL contained in the query request, for dispatching query requests containing URL prefixed by redirect prefix to web page retrieval device 603 and dispatching query requests containing URL prefixed by translation prefix to query translation device 605; web page obtaining device 603 for sending query request to the search engine specified by the URL and receiving the web page requested by the URL; web page modification device 604 for transforming the URLs embedded in the web page received by web page retrieval device 603, i.e. adding translation prefix in front of URL which requires query terms and adding redirect prefix in front of other URLs; query translation device 605 for translating/substituting the query terms in native language of requested URL into/with query terms in the specific language of search engine and adding redirect prefix to the URL; web page generation device for generating new web page and embedding URLs and a segment of script which will be executed automatically when client receives the web page and sends another query request. Furthermore, query translation device 605 also utilizes dictionary 607 and search engine template storage device 608. Dictionary 607 stores abundant dictionary entries, each of them containing a mapping from native language terms to corresponding translations. Search engine template storage device 608 may be referred to as shown in Figures 4 and 5. Web page generation device 606 utilizes a script generator 609 for generating the script executed on the client.

[0095] Figures 7 to 10 are the demonstration with browser when user query is in Chinese using the method and system of this invention.

[0096] Figure 7 shows that user can enter into Yahoo! search engine through another portal site for the query translation server. Of course, it is possible to add prefix of query translation server before Yahoo! URL with other approach instead of using such a portal site.

[0097] Figure 8 shows that user inputs Chinese query term

25

“游

30

戏”

In the homepage of Yahoo! search engine.

[0098] Figure 9 shows the web page returned by Yahoo!, i.e., the query result, as a response of query request for Chinese query term

40

“游戏”

in figure 8 by means of the query translation server of this invention. Browser executes the script and the execution results in another window showing the three English translation choices for

45

“游戏”.

At the same time, user can click on the hypertext entitled "Modify Current Setting" so as to modify the template for Yahoo! search engine.

[0099] Figure 10 shows the browser window after user clicks on the hypertext entitled "Modify Current Setting". It shows the content of template for Yahoo! search engine used by query translation server. User can make appropriate changes.

[0100] By combining the method of this invention with a web page translation service, it is possible to provide complete query translation service. The application is shown in figure 8 and figure 9, that is, user inputs Chinese query terms and receives Chinese query results. On top of that, this invention may also be integrated with speech recognition system to provide service that translates speech query terms into appropriate text query terms.

[0101] Though the above description about the method and system of this invention is in the context of Internet search engine, the design of this invention can be applied to retrieval or query of databases. With the design, a system can be constructed for providing native language query service for users of databases, comprising a plurality of data-

bases, each database having a respective dedicated language, adapted for receiving query requests containing query terms in said dedicated language and returning query results in relation to said query terms, said system characterized by further comprising a query translation server which comprises: a client interface for receiving query requests sent by clients and returning query results to said clients; a query translation apparatus for receiving query terms in user's native language in the query requests received by said client interface into and relating them with query terms in the dedicated language of the database; a query result obtaining apparatus for sending the translated query requests to the databases designated by said query requests and obtaining query results. Similarly, said query terms in native language in query request to the database can be substituted with speech query terms.

[0102] Although certain preferred embodiments of this invention have been described in details with reference to the accompanying drawings, those skilled in the art should appreciate that various changes and modifications may be made therein without departing from the spirit and scope of this invention. Therefore, the scope of this invention is only defined by the Claims.

Claims

1. A method for providing native language query service for Internet users by using a plurality of search engines in the Internet, each search engine having a respective dedicated language, adapted for receiving query requests containing query words of said dedicated language and returning query results in relation to said query words, said method characterized by comprising steps of:

(a) receiving at a site an original query request from one of said Internet users, said original query request containing query words of native language of said user;

(b) selecting a suitable search engine from said plurality of search engines;

(c) translating said query words of native language into query words of dedicated language of said selected search engine;

(d) constructing a new query request directing to said selected search engine based on said original query request and said query words of dedicated language;

(e) sending said new query request to said selected search engine and receiving a returned query result;

(f) sending said query result back to said user as a query result in relation to said original query request.

2. The method according to claim 1, characterized in that said step (b) comprises a step of selecting a search engine designated by an URL in said original query request as the selected search engine.

3. The method according to claim 2, characterized in that said step (c) comprises steps of:

on the basis of the URL in said original query request, retrieving a search engine template matching said URL from a search engine template storage;

translating said query words of native language into query words of a dedicated language defined in said retrieved search engine.

4. The method according to claim 3, characterized in that said step (c) further comprises steps of:

if no search engine template matching said URL is retrieved from said search engine template storage, searching a dedicated language corresponding to said URL from history records in said site based on said URL;

determining positions of query word parameters by using linguistic characteristics of parameter values;

translating query words of native language at said positions into query words of said dedicated language.

5. The method according to claim 3 or 4, characterized in that said step (d) comprises a step of replacing query words of native language in said original query request with query words of said dedicated language so as to form said new query request.

6. A method for providing native language query service for Internet users by using a plurality of search engines in the Internet, each search engine having a respective dedicated language, adapted for receiving query requests containing query words of said dedicated language and returning query results in relation to said query words, said method characterized by comprising steps of:

5

(a) receiving at a site an original query request from said Internet user, said original query request containing an URL requested by said user, said URL having a prefix for designating said site;

10

(b) removing said prefix from said URL;

(c) selecting a suitable search engine from said plurality of search engines;

(d) if said removed prefix is a redirect prefix, performing following steps of:

15

(d1) sending a request containing said URL to said selected search engine and receiving a web page as a response;

(d2) adding a translation prefix before URLs that need query words and a redirect prefix before other URLs in said web page, so as to form a new web page;

20

(e) if said removed prefix is a translation prefix, performing following steps of:

(e1) translating query words of user's native language in parameters of said URL into query words of a dedicated language of said selected search engine;

25

(e2) replacing query words of user's native language in parameters of said URL with query words of said dedicated language;

(e3) adding a redirect prefix before said URL;

30

(e4) generating a new web page, embedding said URL and a Script program in said web page, said Script program enabling a client which receives said new web page to perform a step of automatically sending another original query request based on said URL embedded in said web page;

35

(f) sending said new web page back to said user as a query result in relation to said original query request.

7. The method according to claim 6, characterized in that said step (c) comprises a step of selecting a search engine designated by said URL as said selected search engine.

40

8. The method according to claim 7, characterized in that said step (e1) comprises steps of:

on the basis of said URL, retrieving a search engine template matching said URL from a search engine template storage;

45

translating said query words of native language into query words of a dedicated language defined in said retrieved search engine.

9. The method according to claim 8, characterized in that said step (e1) further comprises steps of:

50

if no search engine template matching said URL is retrieved from said search engine template storage, searching a dedicated language corresponding to said URL from history records in said site based on said URL;

determining positions of query word parameters by using linguistic characteristics of parameter values;

55

translating query words of native language at said positions into query words of said dedicated language.

10. The method according to claim 6, characterized in that said step (e) is

(a) If said removed prefix is a translation prefix, performing following steps of:

(a1) translating a query word of user's native language in parameters of said URL into a plurality of query words of a dedicated language of said selected search engine;

(a2) replacing the query word of user's native language in parameters of said URL with each of said plurality of query words of said dedicated language respectively and forming a plurality of URLs;

(a3) adding a redirect prefix before each of said plurality of URLs;

(a4) setting one of said plurality of URLs as a default URL;

(a5) generating a new web page, embedding said plurality of URLs and a Script program in said web page, said Script program enabling a client which receives said new web page to perform steps of: displaying said plurality of query words of said dedicated language in hyper text format, and automatically sending another original query request based on said default URL embedded in said web page.

11. A system for providing native language query service for Internet users, comprising a plurality of search engines in the Internet, each search engine having a respective dedicated language, adapted for receiving query requests containing query words of said dedicated language and returning query results in relation to said query words, said system characterized by further comprising a query translation server which comprises:

a client interface, for receiving query requests sent by clients and returning query results to said clients;

a request distribution apparatus, for receiving said query requests from said client interface, removing prefixes from requested URLs, and distributing said query requests to different components;

a web page retrieving apparatus, for receiving a query request whose prefix is a redirect prefix from said request distribution apparatus, sending said query request to a search engine designated by an URL and obtaining a requested web page;

a web page modification apparatus, for forming a new web page by adding translation prefixes before URLs that need query words and adding redirect prefixes before other URLs in the obtained web page, and sending said new web page to said client interface;

a query translation apparatus, for receiving a query request whose prefix is a translation prefix from said request distribution apparatus, translating query words of user's native language in the requested URL into and replacing them with query words of a dedicated language of said search engine, and adding a redirect prefix before said URL;

a web page generation apparatus, for generating a new web page, embedding said URL and a Script program in said web page, and sending said new web page to said client interface, said Script program enabling a client which receives said new web page to perform a step of automatically sending another query request based on said URL embedded in said web page.

12. The system according to claim 11, characterized in that said query words of native language are speech query words.

13. A system for providing native language query service for database users, comprising a plurality of databases, each database having a respective dedicated language, adapted for receiving query requests containing query words of said dedicated language and returning query results in relation to said query words, said system characterized by further comprising a query translation server which comprises:

a client interface, for receiving query requests sent by clients and returning query results to said client;

a query translation apparatus, for translating query words of user's native language in the query requests received by said client interface into and replacing them with query words of the dedicated language of the database;

EP 1 072 984 A2

a query result obtaining apparatus, for sending the translated query requests to the databases designated by said query requests and obtaining query results.

- 5 14. The system according to claim 13, characterized in that said query words of native language are speech query words.

10

15

20

25

30

35

40

45

50

55

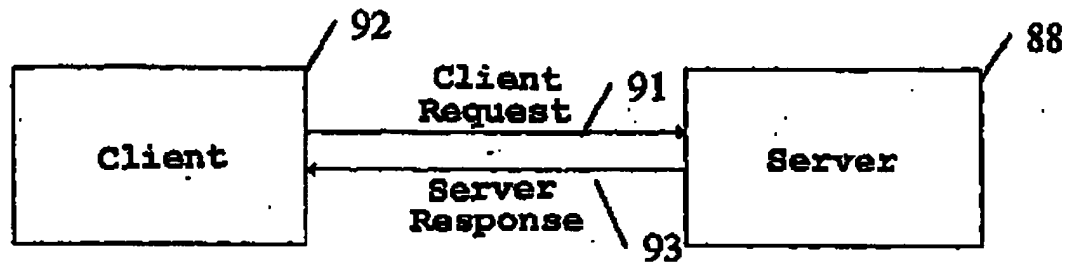


FIG. 1

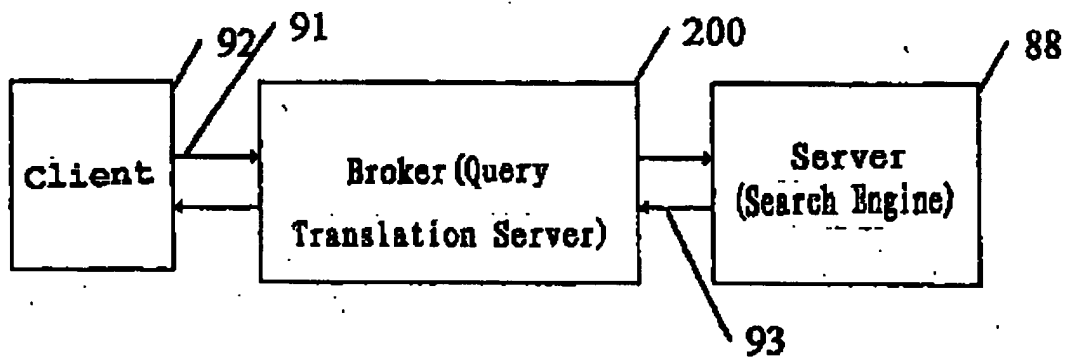
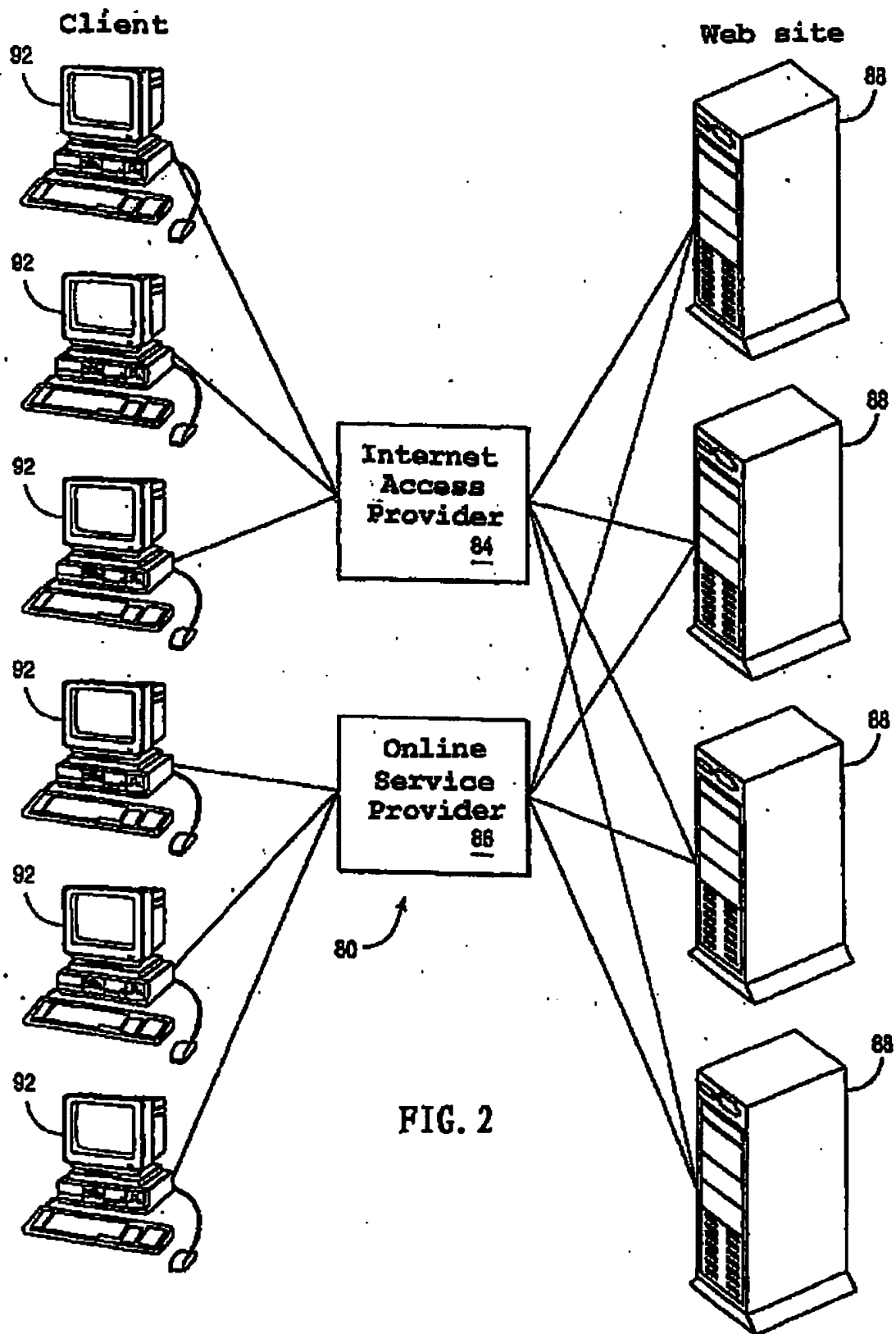


FIG. 3



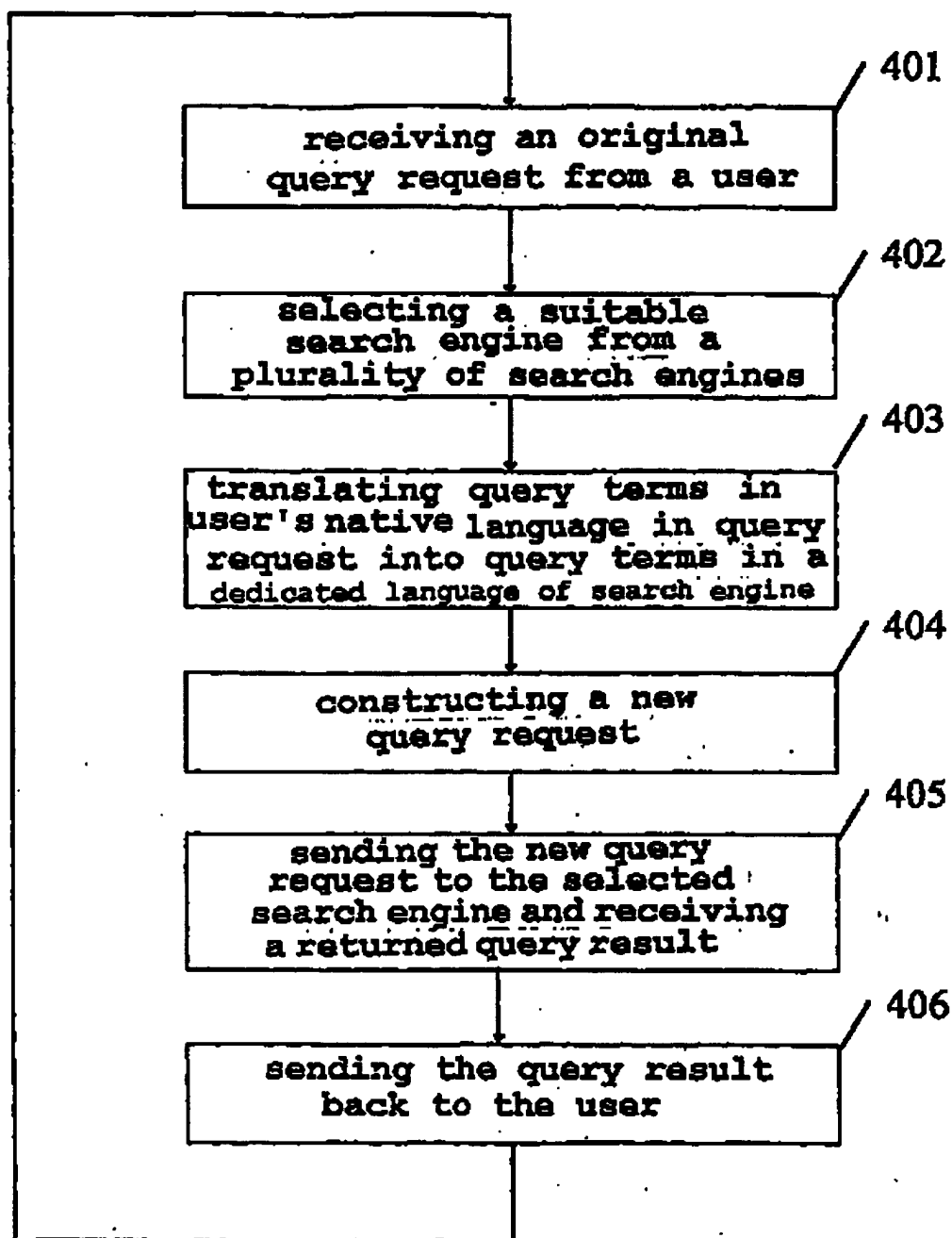


FIG. 4

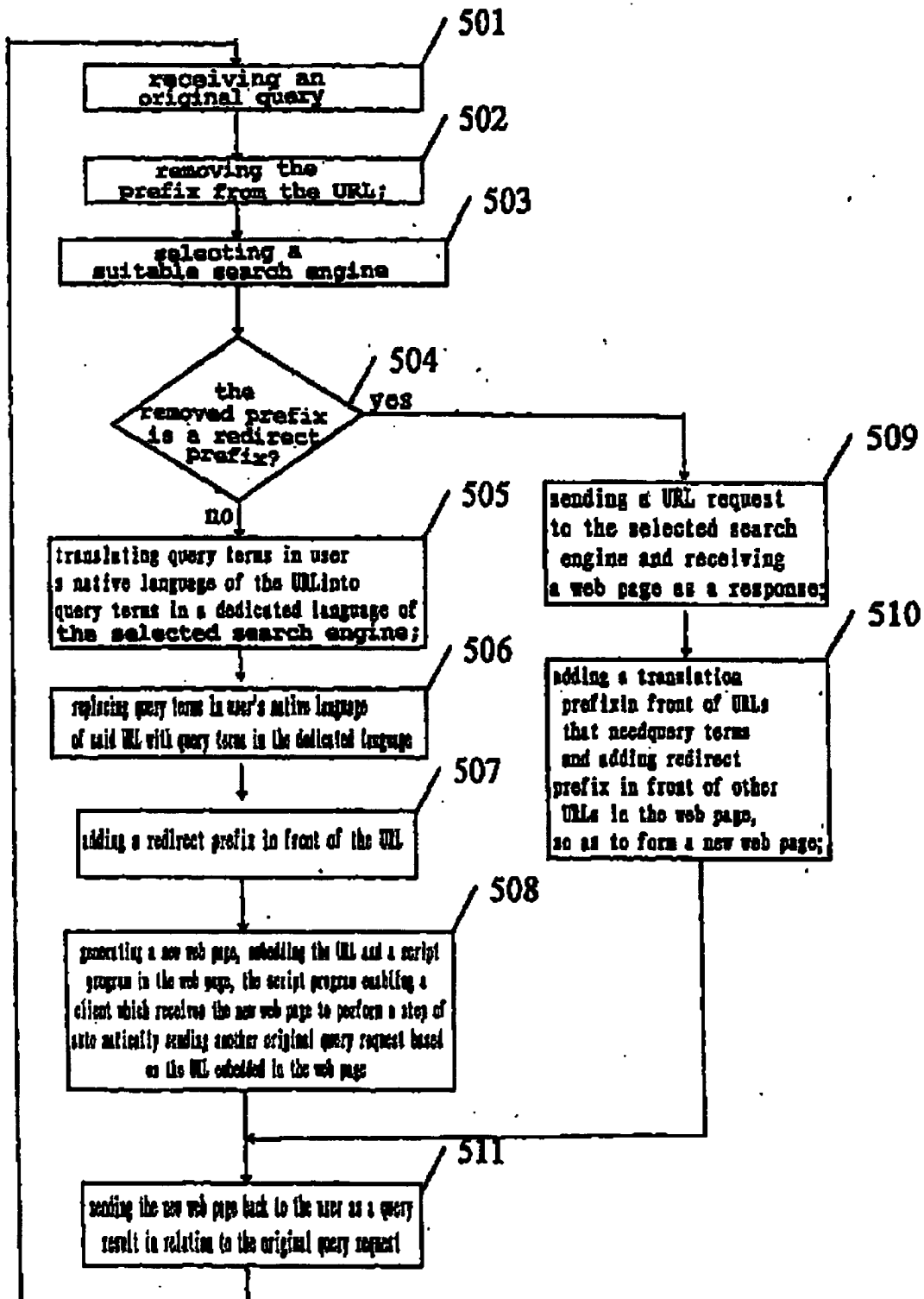
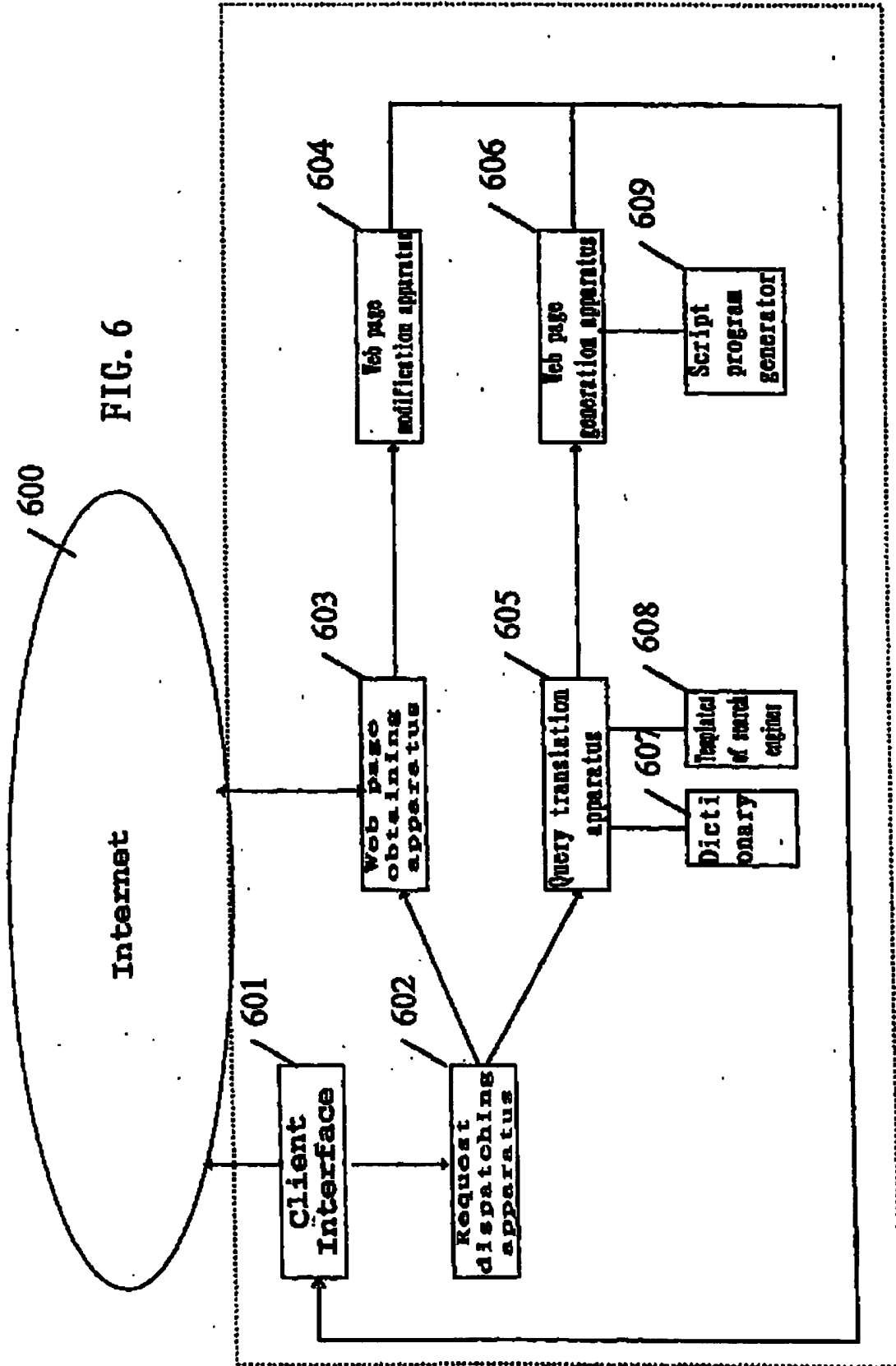


FIG. 5



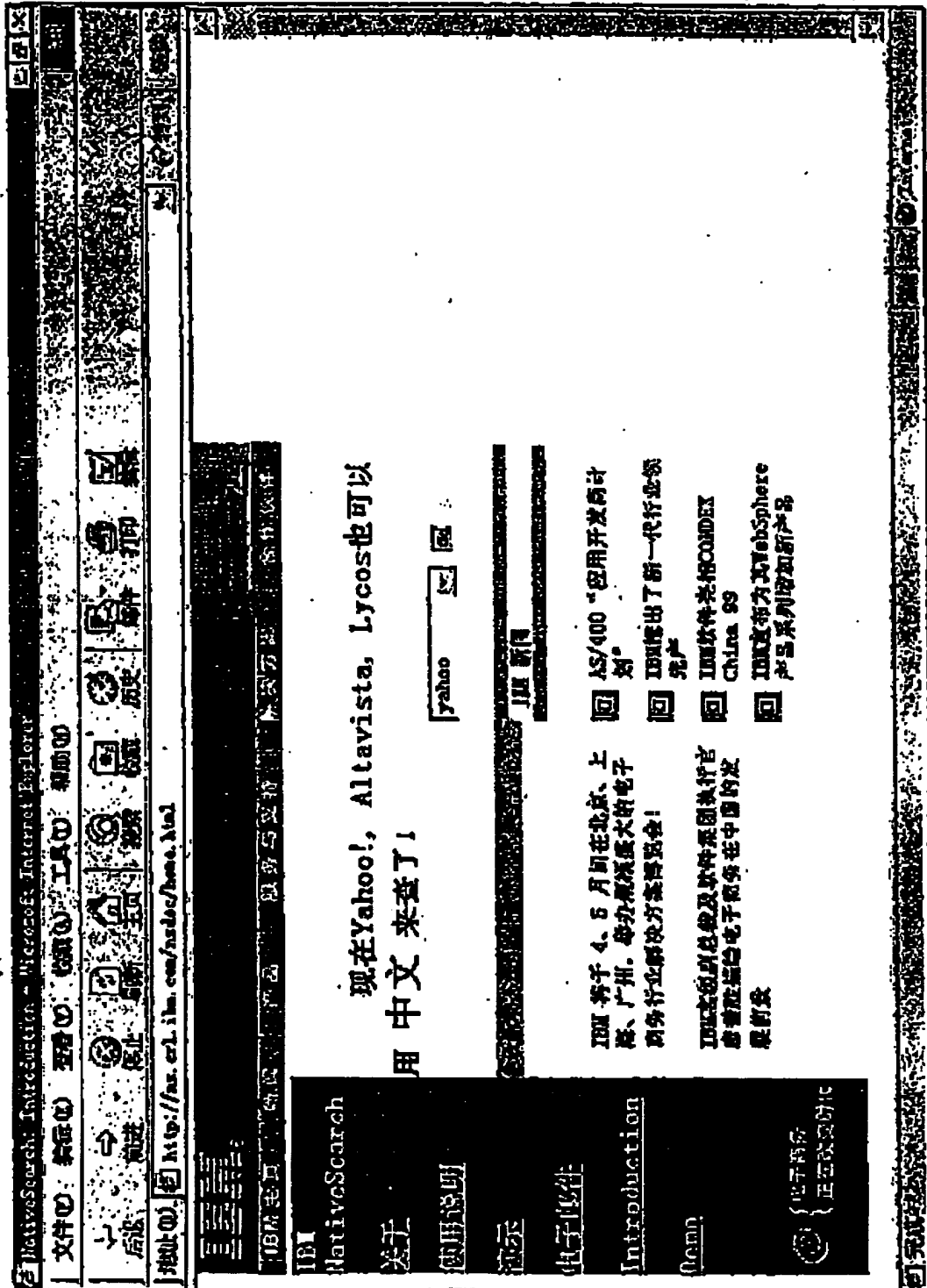


FIG. 7

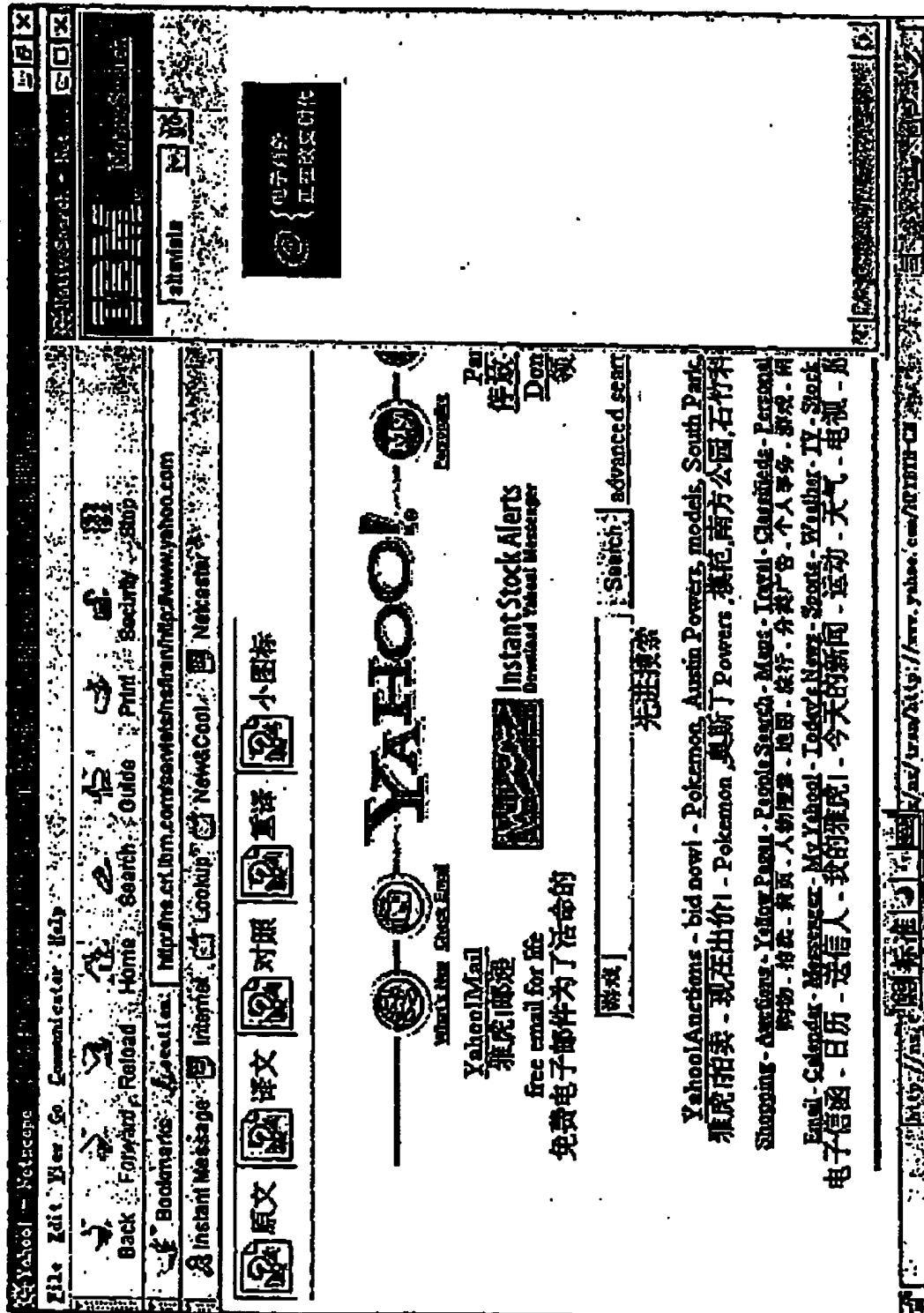


FIG. 8.

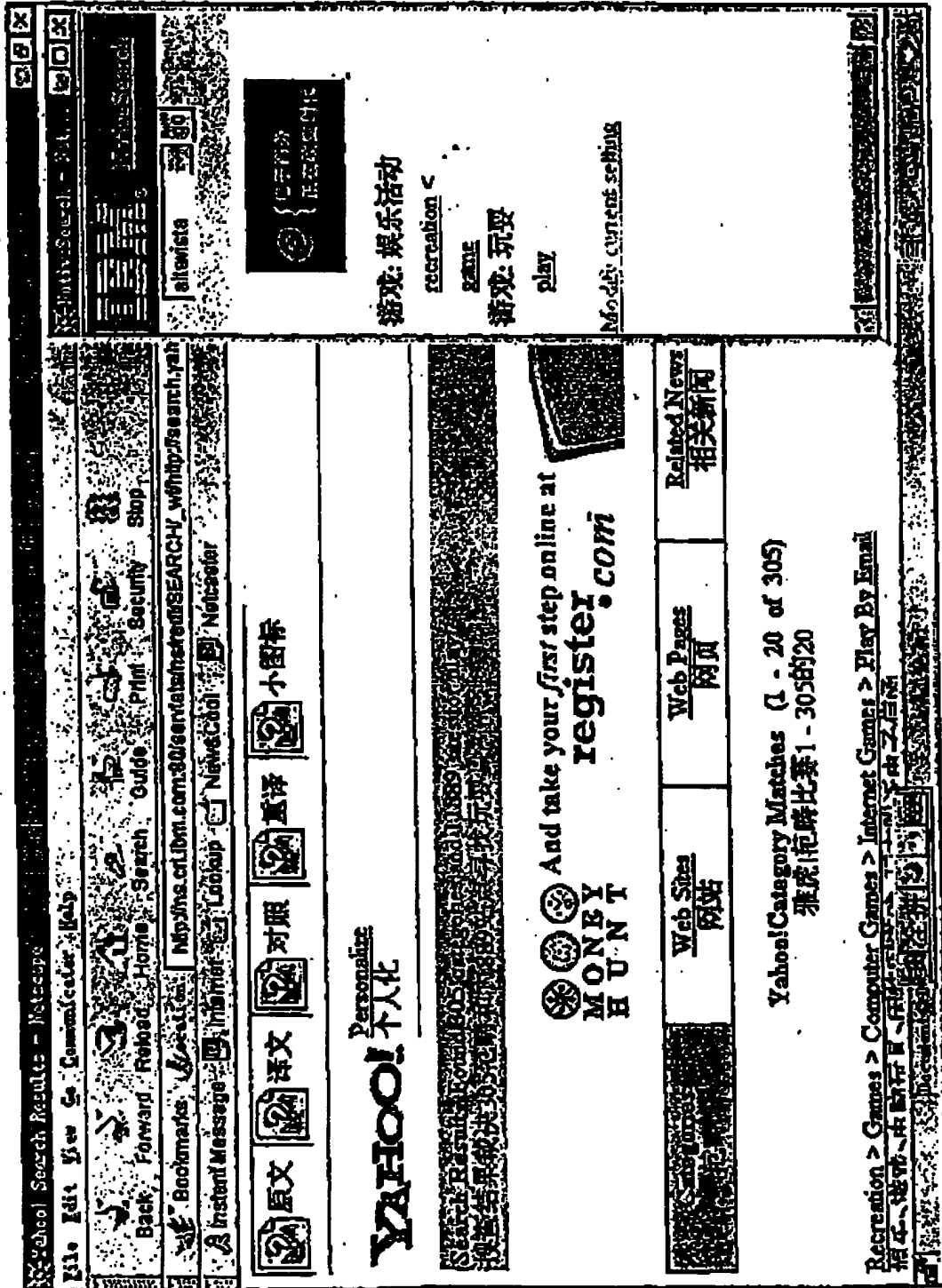


FIG. 9

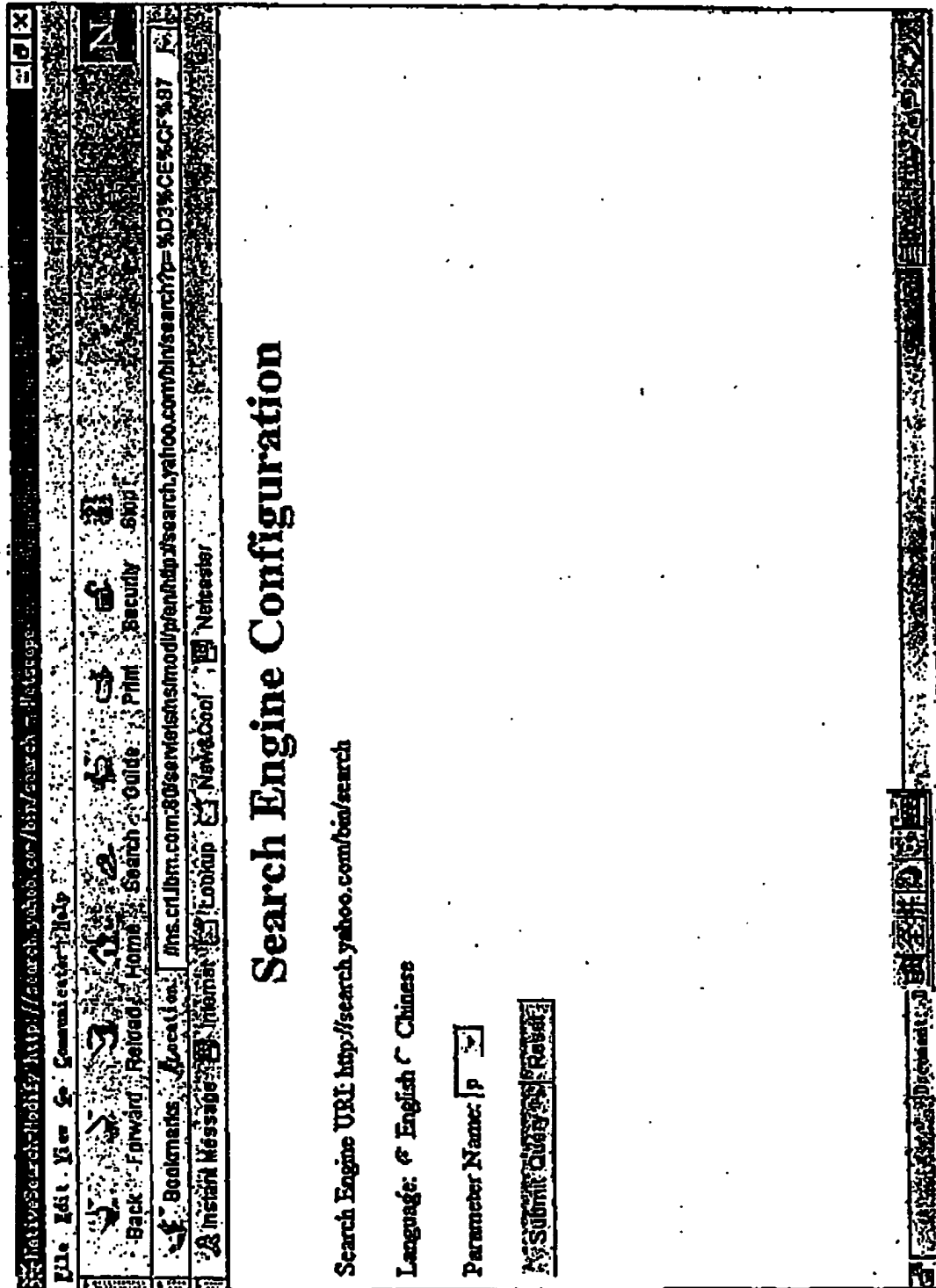


FIG. 10